UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/563,311	12/30/2005	Eiji Murakami	96790P517	6897

8791 7590 01/22/2009

BLAKELY SOKOLOFF TAYLOR & ZAFMAN LLP
1279 OAKMEAD PARKWAY
SUNNYVALE, CA 94085-4040

EXAMINER	
HWA, SHYUE JIUNN	

ART UNIT	PAPER NUMBER
2163	

MAIL DATE	DELIVERY MODE
01/22/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/563,311	MURAKAMI ET AL.	
	Examiner	Art Unit	
	JAMES HWA	2163	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *30 October 2008*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,3-5,8-10,12-14,17 and 18* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,3-5,8-10,12-14,17 and 18* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Applicant has amended claims 1 and 10 in the amendment filed on 10/30/2008. Claims 1, 3-5, 8-10, 12-14, 17 and 18 are pending in this Office Action.

Response to Arguments

2. Applicant's arguments with respect to claims 1, 3-5, 8-10, 12-14, 17 and 18 have been considered but are moot in view of the new ground(s) of rejection.

Applicant argued with regarding 101 issues, the index provides a useful, concrete, and tangible result in the form of an indication of the validity of the edition from the DT matrices. This information is useful for determining whether a hierarchical cluster of documents is valid. Document classification is greatly simplified if the document is classified using the hierarchical clustering and validity index as recited by Claim 1. Examiner respectfully disagrees.

In response to Applicant's argument, claims 1, 3-5 and 8-9, "a sentence classification device comprising an index generation module for making said DT matrix generation module generate DT matrices by using term lists before and after edition by said term list edition module" lack the necessary physical articles or objects to constitute a machine or a manufacture. Consequently, the rejection to claims 1, 3-5 and 8-9 under 35 U.S.C. 101 is maintained.

Claim Rejections - 35 USC § 101

3. Claims 1, 3-5, 8-10, 12-14 and 17-18 are rejected under 35 U.S.C.101 because the language of the claim raises a question as to whether the claim is directed merely to

an abstract idea that is not tied to a technological art, environment or machine which would result in a practice application producing a concrete, useful, and tangible result to form the basis of statutory subject matter under 35 U.S.C 101.

The claims 1, 3-5 and 8-9 lack the necessary physical articles or objects to constitute a machine or a manufacture within the meaning of 35 USC 101. They are clearly not a series of steps or act to be a process nor are they a combination of chemical compounds to be a composition of matter. As such, they fail to fall within a statutory category. They are, at best, functional descriptive material *per se*.

The claims 10, 12-14 and 17-18 recite the mental steps that do not tied to statutory class (such as a particular apparatus). In particularly, a method claim would not qualify as a statutory process would be a claim that recited purely metal steps. Thus, to qualify as a 101 statutory process, the claim should positively recite the other statutory class (the thing or product) to which it is tied, for example by identifying the apparatus that accomplishes the method steps or positively recite the subject matter that is being transformed, for example by identifying the material that is being changed to a different state.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

4. Claims 1, 3-5, 8-10, 12-14 and 17-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tokuda et al. (US Patent No. 7,024,400 B2, hereinafter "Tokuda") in view of Handa et al. (US Patent No. 6,067,259 A, hereinafter "Handa"), Agrawal et al. (US Patent Application No. 2001/0037324 A1, hereinafter "Agrawal") and Bent et al. (US Patent Application No. 2004/0205457 A1, hereinafter "Bent").

As to claims 1 and 10, Tokuda teaches the claimed limitations:

"A sentence classification device" as document classification is important not only in office document processing but also in implementing an efficient information retrieval system (column 1, lines 13-15).

"A term list having a plurality of terms each comprising not less than one word" as a term is defined as a word or a phrase that appears in at least two documents (column 4, lines 5-6).

"a DT matrix generation module for generating a DT matrix two-dimensionally expressing a relationship between each document contained in a document set and

said each term" as the term by document matrix of the original documents (column 9, lines 23; see also table 1).

"a DT matrix transformation module for generating a transformed DT matrix having clusters having blocks of associated documents by transforming the DT matrix obtained by said DT matrix generation module" as exploiting the singular vector decomposition method, the major left singular vectors associated with the largest singular values are selected as a major vector space called an intra-DLSI space, or an I-DLSI space (column 3, lines 2-5). The extra-DLSI space or the E-DLSI space can similarly be obtained by setting up a differential term by extra-document matrix where each column of the matrix denotes a differential document vector between the document vector and the centroid vector of the cluster, which does not include the document. The extra-DLSI space may then be constructed by the major left singular vectors associated with the largest singular values (column 3, lines 18-25).

"a classification generation module for generating classifications associated with the document set on the basis of a relationship between each cluster on the transformed DT matrix obtained by said DT matrix transformation module and said each document classified according to the clusters; wherein the classification generation module comprises a virtual representative document generation module for generating a virtual representative document, for each cluster on a transformed DT matrix, from a term of each document belonging to the cluster" as an automatic document classification method using a DLSI space-based classifier operating on a computer as a computerized classifier to classify a document in accordance with clusters in a

database, comprising the steps of: a) setting up, by said computerized classifier, a document vector by generating terms as well as frequencies of occurrence of said terms in the document, so that a normalized document vector N is obtained for the document (claim 2). Given a new document to be classified, a best candidate cluster to be recalled from the clusters can be selected from among those clusters having the highest probabilities of being the given differential intra-document vector (column 3, lines 10-13). The differences in word usage between the document and a cluster's centroid vector, the differential document vector is capable of capturing the relation between the particular document and the cluster (Column 2, lines 41-46).

Tokuda does not explicitly teach the claimed limitation “DT matrix generation module on the basis of a DM decomposition method in a graph theory; at each DT matrix transformation, said DM decomposition method used to hierarchically cluster documents by setting said DT matrix generated by said DT matrix generation module”.

Handa teaches the information on the positions of faulty elements in the whole memory including the spare lines and limbos is expressed in the form of a bipartite graph to calculate the maximum matching thereof, whereby it can be decided in a short time that the faulty elements are unrepairable by the number of spare lines provided. Further, by the use of the DM decomposition calculated from the maximum matching, the optimum or quasi-optimum relationship of correspondence between the faulty lines and the spare lines can be determined in a short time. In particular, in cases such as, e.g., in case faulty elements take place in a uniformly dispersed state or in case many of the faults left after a line fail has been repaired are in the form of single faults, it can be

expected to calculate the relationship of correspondence realized in the minimum number (column 8, lines 25-37).

The DM decomposition generating unit evaluates a DM decomposition by the use of the data of the maximum matching thus obtained and the contracted graph thereof. The process ranging from the line fail detection unit to the repair solution generating unit is an operation directed to the extended matrix B. In the repair solution output unit, the operation of calculating, for the first time, the relationship between the faulty lines in the original memory block matrix A and the original spare lines by discriminating between them (column 21, lines 22-60).

Tokuda does not explicitly teach the claimed limitation “a large classification generation module for generating a large classification of documents from each document in a bottom-up manner by repeatedly performing, at each DT matrix transformation, used to hierarchically cluster documents by setting said DT matrix generated by said DT matrix generation module in an initial state, causing said virtual representative document generation module to generate a virtual representative document for each cluster on a transformed DT matrix generated from the DT matrix by said DT matrix transformation module, generating a new DT matrix used for next hierarchical clustering processing by adding the virtual representative document to the transformed DT matrix and deleting documents belonging to the cluster of the virtual representative document from the transformed DT matrix, and outputting, for said each cluster, information associated with the documents constituting the cluster as large

classification data; and generating and outputting an index indicating validity of the edition from the DT matrices ".

Agrawal teaches for organizing a large text database into a hierarchy of topics and for maintaining this organization as documents are added and deleted and as the topic hierarchy changes. The hierarchical technique can handle millions of documents and tens of thousands of topics. A resulting taxonomy and path enhanced retrieval system (TAPER) is used to generate context-dependent document indexing terms. The topic paths are used, in addition to keywords, for better focused searching and browsing of the text database (abstract). For such classifiers, feature sets larger than 100 are considered extremely large. Document classification may require more than 50,000. Singular value decomposition on the term-document matrix has been found to cluster semantically related documents together even if they do not share keywords (page 2, paragraph 0019-0021).

The feature set changes by context as the classification process proceeds down the taxonomy. As a result, jargon common to lower nodes of the taxonomy are filtered out and the classification accuracy remains high in spite of the reduction in the number of terms and candidate classes inspected (page 3, paragraph 0029). Each document in the database has been pre-classified. The user may then enter a command through the user input device to cause the system to select at least one of the displayed sub-topics. This process is repeated as necessary to refine the query topic until the user's information need is satisfied (page 5, paragraph 0084).

A parent class inherits, in an additive fashion, the statistics of its children, since each training document generates rows for each topic node from the assigned topic up to the root (page 13, paragraph 0204).

Although Tokuda teaches preprocessing documents using said computer to distinguish terms of a word and a noun phrase from stop words; constructing system terms by setting up a term list as well as global weights using said computer (claim 1). The method includes the setting up of a differential latent semantics index (DLSI) space-based classifier to be stored in computer storage and the use of such classifier by a computer to evaluate the possibility of a document belonging to a given cluster using a posteriori probability function (abstract).

Tokuda does not explicitly teach the claimed limitation "a term list edition module for adding or deleting an arbitrary term with respect to the term list; and an index generation module for making said DT matrix generation module generate DT matrices by using term lists before and after edition by said term list edition module".

Bent teaches an initial document by term matrix is formed, each document being represented by a respective M dimensional vector, where M represents the number of terms or words in a predetermined domain of documents. The techniques of text mining currently include the automatic indexing of documents, extraction of key words and terms, grouping/clustering of similar documents, categorising of documents into pre-defined categories and document summarization (page 1, paragraph 0010-0011).

TextFormatter reads both the textual document in the document set and the term list generated (page 4, paragraph 0060; see also element 305 of figure 3). The text from

the document is read in and tokenised into sentences. Sentences again are tokenised into words. Now the sentences have to be checked for terms that have an entry in the hashtable. Since it is possible that words which are part of a composed term occur as single words as well, it is necessary to check a sentence backwards. That is, firstly the hashtable is searched for a test string which consists of the whole sentence. When no valid entry is found one word is removed from the end of the test string and the hashtable is searched again. This is repeated until either a valid entry was found or only a single word remains.

To be admitted as a column of the term-sentence matrix, a term must occur in the sentences of the document set more often than a minimum frequency, whereby a user or administrator may determine the minimum frequency. For instance, it is illogical to add terms to the matrix that occur only once, as the objective is to find clusters of sentences which have terms in common. Next, the document vector is searched for all occurrences of term #1 of the term vector. If the term occurs at least as often as the specified minimum frequency, it remains in the term vector and if the term occurs less often, it is removed. Since actor occurs only once in the document vector, the term is deleted from the head of the term vector (page 4, paragraph 0060-0069).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Handa, Agrawal and Bent before him/her, to modify Tokuda using DM decomposition method to hierarchically cluster documents because that would provide method and device for repairing arrays with redundancy, wherein it can be decided in a short time and

optimum solution concerning the way of using the spare lines as taught by Handa (column 2, lines 45-54).

As to claims 3 and 12, Tokuda teaches the claimed limitations:

"Characterized by further comprising label generation module for outputting each term strongly connected to each document belonging to said arbitrary cluster as a label indicating a classification of the cluster" as a new efficient supervised document classification procedure introduced, whereby learning from a given number of labeled documents preclassified into a finite number of appropriate clusters in the database, the classifier developed will select and classify any of new documents introduced into an appropriate cluster within the classification stage (column 2, lines 21-25).

As to claims 4 and 13, although Tokuda teaches the extra-DLSI space, or the E-DLSI space can similarly be obtained by setting up a differential term by extra-document matrix where each column of the matrix denotes a differential document vector between the document vector and the centroid vector of the cluster which does not include the document (column 3, lines 18-23).

Tokuda does not explicitly teach the claimed limitation "Characterized by further comprising document organization module for sequentially outputting documents belonging to said arbitrary cluster or all documents in an arrangement order of the documents in the transformed DT matrix".

Agrawal teaches given k*(c), the sorted Fisher table is scanned while copying the first k*(c) rows for the run corresponding to class c to an output table and discarding the remaining terms. This involves completely sequential IO (page 12, paragraph 0187).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Handa, Agrawal and Bent before him/her, to modify Tokuda the document organization because that would improve the document search performance include speed and accuracy as taught by Agrawal (page 14, paragraph 0216).

As to claims 5 and 14, Tokuda teaches the claimed limitations:

"Characterized by further comprising summary generation module for outputting, as a summary of said arbitrary document, a sentence of sentences constituting the document which contains a term strongly connected to the document" as the setting up of a DLSI space-based classifier is summarized. Documents are preprocessed, to identify and distinguish terms, either of the word or noun phrase, from stop words. System terms are then constructed, by setting up the term list as well as the global weights. The process continues with normalization of the document vectors, of all the collected documents, as well as the centroid vectors of each cluster. Following document vector normalization, the differential term by document matrices may be constructed by intra-document or extra-document construction (column 7, lines 24-34).

As to claims 8 and 17, Tokuda does not explicitly teach the claimed limitation "characterized in that said large classification generation module terminates repetition of the clustering processing when no cluster is obtained from the transformed DT matrix in the clustering processing".

Agrawal teaches each of the other second level topics may be divided at the third level to further topics. Also, in a similar fashion, further levels under the third level may be included in the topic hierarchy, or taxonomy. The final level of each path in the taxonomy terminates at a terminal or leaf node (page 6, paragraph 0087). Large sub-trees in the topic tree can be eliminated forthwith if the score of the root of those sub-trees are very poor (page 8, paragraph 0131).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Handa, Agrawal and Bent before him/her, to modify Tokuda terminates repetition of the clustering processing because that would provide a means for designing vastly enhanced searching, browsing and filtering systems as taught by Agrawal (page 1, paragraph 0009).

As to claims 9 and 18, Tokuda teaches the claimed limitations:

"Characterized by further comprising large classification label generation module for, if a virtual representative document is contained in a given cluster of clusters obtained by the clustering processing" as a new efficient supervised document classification procedure, whereby learning from a given number of labeled documents preclassified into a finite number of appropriate clusters in the database, the classifier

developed will select and classify any of new documents introduced into an appropriate cluster within the classification stage (column 2, lines 22-28).

Tokuda does not explicitly teach the claimed limitation "generating a label of the cluster on which the virtual representative document is based from a term strongly connected to the virtual representative document".

Agrawal teaches that with reference to the hierarchy represented, statistics are calculated for the science node, based on the terms in all of the documents from the collection set that are classified in classes represented by nodes below the science node. Including the nodes labeled biology, chemistry, electronics, and all children nodes of those nodes (page 6, paragraph 0093). Large sub-trees in the topic tree can be eliminated forthwith if the score of the root of those sub-trees are very poor. Text database population is not the only application of fast multi-level classification. With increasing connectivity, it will be inevitable that some searches will go out to remote text servers and retrieve results that must then be classified in real time (page 8, paragraph 0131).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Handa, Agrawal and Bent before him/her, to modify Tokuda strongly connected to the virtual representative document because that would provide a system which is sufficiently fast as taught by Agrawal (page 2, paragraph 0025).

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to James Hwa whose telephone number is 571-270-1285. The examiner can normally be reached on 8:00 – 5:00. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Wong can be reached on 571-272-1834. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only, for more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the PAIR system contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

01/08/2009

/James Hwa/
Examiner, Art Unit 2163

/Cam Y Truong/
Primary Examiner, Art Unit 2169